

# SIMULATED ANNEALING AND BAYESIAN POSTERIOR DISTRIBUTION ANALYSIS APPLIED TO SPECTRAL EMISSION LINE FITTING.

Jack Ireland

*ADNET Systems, Inc., NASA's Goddard Spaceflight Center, Mail Code 612.5,  
Greenbelt, MD 20771, USA (ireland@grace.nascom.nasa.gov)*

Received 17 November 2006; accepted 16 April 2007

**Abstract.** Spectral-line fitting problems are extremely common in all remote-sensing disciplines, solar physics included. Spectra in solar physics are frequently parameterized using a model for the background and the emission lines, and various computational techniques are used to find values to the parameters given the data. However, the most commonly-used techniques, such as least-squares fitting, are highly dependent on the initial parameter values used and are therefore biased. In addition, these routines occasionally fail due to ill-conditioning. Simulated annealing and Bayesian posterior distribution analysis offer different approaches to finding parameter values through a directed, but random, search of the parameter space. The algorithms proposed here easily incorporate any other available information about the emission spectrum, which is shown to improve the fit. Example algorithms are given and their performance is compared to a least-squares algorithm for test data – a single emission line, a blended line, and very low signal-to-noise ratio data. It is found that the algorithms proposed here perform at least as well or better than standard fitting practices, particularly in the case of very low signal-to-noise ratio data. A hybrid simulated annealing and Bayesian posterior algorithm is used to analyze a Mg X line contaminated by an O IV triplet, as observed by the Coronal Diagnostic Spectrometer (CDS) onboard SOHO. The benefits of these algorithms are also discussed.

**Keywords:** Methods

## 1. Introduction

Solar spectra, comprising continuum emission, line emission (or absorption) background emission and noise, can provide useful and detailed information on the physical state (for example, plasma density, bulk plasma flow, turbulent flow) of the emitting volume. To extract such information, one typically assume a profile for each line thought to be present and a varying background signal. This model spectrum is then fitted to the data. However, the fitting procedure used can make a difference to the information extracted. A study by Brynildsen (1994) of the Gaussian fitting routines used in the support software to the CDS and SUMER (Solar Ultraviolet Measurements of Emitted Radiation) instruments onboard SOHO makes this amply clear. The report con-

cludes that the fitting techniques used – all dependent on some form of user-supplied initial guesses (whether automated or interactively set) and deterministic algorithms – work well when there is one Gaussian component present, but become dependent on the initial guess when two or three components are present.

Unfortunately, there are many portions of the solar spectrum that contain multiple and overlapping lines, and so generating reliable parameter values becomes much more problematic. For such spectra, other approaches can be tried: for example, McIntosh *et. al.* (1998) examined the applicability of genetic algorithms to spectral-line fitting problems. It was found that the dependence on initial guess can be largely eliminated, at the expense of much longer computation times. A genetic algorithm is an example of a random, but directed, search algorithm: this algorithm relies on an analogy with the process of evolution, finding local minima in the search space, whilst also permitting the possibility to jump to other smaller local minima (including the global minimum).

Line-fitting algorithms find estimates to line model parameter values by first calculating a function that describes the difference between the data and the model, known as the cost function. The parameter values are varied until a point in the parameter search space is found where the cost function is minimized. The commonly used least-squares approach uses a cost function based on the (weighted) sum of the square of differences between the data and the model; new parameter values are found by using the gradient of the cost function to move to parameter-space areas that give lower values of the cost function.

Simulated annealing (SA) and Bayesian posterior distribution analysis via Markov chain Monte Carlo (MCMC) algorithms offer different ways to traverse the parameter space to look for minima in the cost function used to find a fit.

These algorithms bring two advantages to the analysis of spectra. As will be shown in Section 5, routines such as IDL CURVEFIT (a least-squares based algorithm) give final results which are highly sensitive to the starting point chosen. Simulated annealing requires no special starting point in order to seed the fitting process, and so the final answer is independent of the user's bias in initial choice. Bayesian posterior distribution analysis allows the user to specify the statistics of the observed data, and to include any prior information on the parameters as part of the process in finding the best explanation of the data in terms of the given spectral model. This frees the user from considering only Gaussian statistics, implicit in IDL CURVEFIT and in most other fitting routines. This is important in low count regimes where the Gaussian approximation breaks down and Poisson statistics must be

used. In addition, if the user has information about the observation, for example, a minimal instrumental line width, or an expectation to observe Doppler shifts, then this too can be included as part of the process in finding the best parameter estimates.

This paper applies these two algorithms to test data in order to examine their utility in fitting model spectra to data. Section 2.1 describes the development of a simulated-annealing algorithm for spectral-line fitting, whilst Section 2.2 describes a Bayesian posterior distribution analysis. Section 4 takes both algorithms and applies them to the analysis of a Mg X line observed by SOHO/CDS. Section 5 compares the effect of using different assumptions of noise in the data for very low signal-to-noise ratio spectra. Finally, Section 6 examines the case for using these algorithms in routine data analysis.

## 2. Fitting Algorithms

### 2.1. SPECTRAL MODEL FITTING BY SIMULATED ANNEALING

Simulated annealing was introduced by Kirkpatrick *et al* (1983) as a solution method to the travelling salesperson problem. It is a function minimization strategy based on an analogy with the creation of crystals from their melts. Since its introduction, it has been applied to many more problems, including the minimization of multi-modal function with continuous variables (Corana *et al.* 1987), and more recently, to finding separatrices in 3D magnetic fields (Beveridge 2006).

In annealing, materials are reheated and slowly cooled in order to reduce the number of defects in the final crystal structure (defects in the structure trap energy within the crystal). Simulated annealing makes analogies between the energy contained in the structure ( $E$ , the function to be minimized), the configurations ( $C$ ) of the crystal structure ( $E$  is a function of  $C$ ) and temperature ( $T$ , an annealing schedule controlling the convergence of the algorithm to a final configuration). A generalised simulated annealing algorithm is presented in Figure 1. At large values of  $T$ , the algorithm behaves like a random search since almost all  $C_{new}$  have energies  $E_{new}$  that satisfy the Metropolis criterion (Metropolis *et al.* 1953). As  $T$  is lowered (analogous to cooling), more and more  $C_{new}$  become unacceptable, with only a few  $C_{new}$  satisfying the Metropolis criterion (highlighted by asterisks in Figure 1). At this stage the algorithm behaves like a localised search with hill climbing to allow it to jump out of local minima. Finally, as  $T$  becomes yet smaller, the chances of moving significantly away from the current local minimum become very small, and the routine has, for

**Procedure Simulated Annealing**

```

begin
  generate initial configuration  $C_{current}$ 
  set initial temperature  $T_0$ 
   $t = 1$ 
  repeat
    repeat
      generate new configuration  $C_{new}$  from  $C_{current}$ 
      evaluate  $E_{new}, E_{current}$ 
      * if ( $E_{new} \leq E_{current}$ ) then begin
      *   accept  $C_{new}$ : becomes  $C_{current}$  for next iteration
      * endif else begin
      *   pick a number  $p$  chosen uniformly randomly in the range  $[0, 1)$ 
      *   if ( $\exp\left\{-\left[\frac{E_{new}-E_{current}}{T(t)}\right]\right\} > p$ ) then begin
      *     accept  $C_{new}$ : becomes  $C_{current}$  for next iteration
      *   endif
      * endelse
    until (equilibrium condition reached)
     $T_{t+1} = \tau\{T_t\}$ 
     $t = t + 1$ 
  until (procedure termination condition satisfied)
  final configuration  $C_{final}$ 
end

```

Figure 1. Scheme for general simulated annealing algorithm (pseudo-code adapted from de Vicente *et al.*, 2003).

most practical purposes, converged to an optimal solution. The general algorithm above conceals a number of choices that are left to the user. The algorithm says nothing on how a new configuration is generated from a previous one, how a configuration is to be represented, a suitable choice of energy function to minimize, or how the temperature should be changed.

Spectra of the form

$$F_i = \mu + \sum_{j=1}^{N_g} A_j \exp \left[ \frac{(x_i - c_j)^2}{2\sigma_j^2} \right] \quad (1)$$

are considered, where there are  $N_d$  data values  $\mathbf{D} = \{D_1, D_2, \dots, D_{N_d}\}$  observed at  $\mathbf{x} = \{x_1, x_2, \dots, x_{N_d}\}$ . A configuration  $C$  is formed from  $\mu, A_1, c_1, \sigma_1, \dots, A_{N_g}, c_{N_g}, \sigma_{N_g}$  ( $1 + 3N_g$  variables), where  $N_g$  is the number of Gaussians used to fit the data. The energy function is chosen to

permit easy comparison with least-squares fitting algorithms; therefore,

$$E = \sum_{i=1}^{N_d} \frac{(F_i - D_i)^2}{W_i^2} \quad (2)$$

for some set of weights ( $W_i$ ). A new variable  $v_{new}$  is calculated from the old  $v_{old}$  through

$$v_{new} = v_{old} + \delta v \quad (3)$$

where  $\delta v = \sqrt{\Upsilon(v)}N(0,1)$  is recalculated until  $v_{min} \leq v_{new} \leq v_{max}$  for upper and lower limits  $v_{max}$  and  $v_{min}$  respectively. The step size ( $\delta v$ ) is calculated from the square root of the Cramer-Rao lower bound  $\Upsilon(v)$  (Ireland, 2005) for each variable. For each variable  $v$ ,  $\Upsilon(v)$  is a lower limit to the variance for a measurement of the variable, and so its square root generates a natural scale size of the search space around the estimated value. It is expected (since the search space is relatively smooth) that better solutions lie close to solutions that are already good and so normally distributed step sizes are used.

The temperature at time  $t$  (the annealing schedule) is given by  $T_t = T_0\delta^t$  for some  $0 \leq \delta < 1$  and initial temperature  $T_0$  (Corana *et al.* 1987 also use this form of temperature dependence in their study of minimizing multimodal functions of continuous variables). Further details of the algorithm initialization are problem-dependent and are given in Section 3.

## 2.2. BAYESIAN POSTERIOR DISTRIBUTION ANALYSIS

Recent developments in spectroscopic analysis have used techniques borrowed from Bayesian statistics and Markov chain Monte Carlo methods. van Dyk *et al.*, (2001) implement ideas from these areas to analyze low-count spectroscopic stellar data. (Kashyap and Drake, 1998) use a Markov chain Monte Carlo based code to reconstruct differential emission distributions from data observed by the Solar Extreme Ultraviolet Rocket Telescope and Spectrograph (Brosius *et al.* 1996). Both these papers take advantage prior information (in the form of probability distributions) into the determination of the parameter values. This results in a posterior distribution encapsulating our knowledge of the model – sampling from this distribution is achieved using a Markov chain Monte Carlo technique.

Let  $C$  be the set of parameters of interest. Then the probability distribution of the parameters given the observed data  $D$

$$p(C|D, I) = \frac{p(D|C, I)p(C|I)}{p(D|I)} \quad (4)$$

where  $I$  represents any initial information known before  $D$  is observed. The quantity  $p(C|I)$  is called the *prior distribution*; it represents prior knowledge of the parameters  $C$  given the information  $I$ . The *sampling distribution* or *likelihood*  $p(D|C, I)$  is the likelihood of the data  $D$  given the model parameters. The left-hand side of Equation (4) is called the *posterior distribution*.

### 2.2.1. Bayesian Posterior Distribution Analysis Via Markov Chain Monte Carlo Methods: The Metropolis-Hastings Algorithm

Having written down the posterior distribution, it remains to draw samples from this probability distribution. Samples can be generated by any process that draws from the target distribution  $p(D|C, I)$  in the correct proportions. Markov chain Monte Carlo methods (MCMC) algorithms generate the samples by constructing a weighted random walk through the parameter space such that the probability of being in a region of the space is proportional to the posterior distribution density for that region. Hence the proportional samples of the posterior distribution are recreated.

The random walk is implemented using a Markov Chain; a sample  $C_{t+1}$  depends on the previous sample  $C_t$  via the transition probability  $p(C_{t+1}|C_t)$ . The transition probability is also assumed to be time independent. After an initial “burn-in” period, the transition probability  $p(C_{t+1}|C_t)$  generates samples  $X$  with the same probability density as the posterior distribution  $p(C|D, I)$ .

Figure 2 is a pseudo-code version of a Metropolis-Hastings Markov chain Monte Carlo algorithm. There are two steps in implementing this. The first step involves picking a proposed value - ( $Y$ ) - that *may* be accepted as the next value  $C_t$ . The value of ( $Y$ ) is chosen from a proposal distribution  $q(Y|C_t)$  which is easy to calculate. The second step is to decide whether the candidate  $Y$  is to be accepted as the value of  $C_{t+1}$ . This is done by calculating

$$r = \frac{p(Y|D, I)q(C_t|Y)}{p(C_t|D, I)q(Y|C_t)}$$

which is known as the Metropolis ratio. If  $r \geq 1$ , then  $Y$  is accepted as the value of  $C_{t+1}$ . If  $r \leq 1$ , then a random number  $U$  is drawn from the uniform distribution between 0 and 1. If  $U \leq r$  then  $Y$  is accepted as the value to  $C_{t+1}$ . If not,  $C_{t+1} = C_t$ . The Metropolis-Hastings algorithm generates samples ( $X$ ) which converge on the target posterior distribution  $p(D|X, I)$  for a wide range of proposal distributions  $q(Y|X)$  (for more details on a proof of convergence see Gregory, 2005).

**Procedure Metropolis-Hastings**

```

begin
  initialize  $C_0$ 
  set  $t = 0$ 
  repeat
    obtain a new sample  $Y$  from  $q(Y|C_t)$ 
    calculate acceptance probability  $r = \frac{p(Y|D,I)q(C_t|Y)}{p(C_t|D,I)q(Y|C_t)}$ 
    if  $r \geq 1$  then begin
       $C_{t+1} = Y$ 
    endif else begin
      pick a number  $U$  chosen uniformly randomly in the range  $[0, 1)$ 
      if  $U \leq r$  then begin
         $C_{t+1} = Y$ 
      endif else begin
         $C_{t+1} = C_t$ 
      endelse
    endelse
    increment  $t$ 
  until (procedure termination condition satisfied)
  Distribution samples are  $C$  from  $t_{\text{burn-in}}$  to  $t_{\text{end}}$ .
end

```

Figure 2. Metropolis-Hastings Markov chain Monte Carlo method pseudo-code (adapted from Gregory, 2005).

### 2.2.2. Algorithm Details

The posterior distribution defined in Equation (4) is proportional to the likelihood and the priors. The likelihood for spectral data is given by

$$p(D|X, I) = \prod_{i=1}^{N_d} \frac{F_i^{D_i}}{D_i!} e^{-F_i} \quad (5)$$

where  $F_i$  depends on the model variables  $C$ . Priors for the emission variables depend on the variable type. Both Gaussian amplitude and background (Gregory, 2005) are given Jeffrey's priors (for example,  $p(A) = 1/[A \ln(\max(A)/\min(A))]$ ), whilst line position and width are given non-informative priors (for example  $p(\sigma) = 1/[\max(\sigma) - \min(\sigma)]$ ). The maximum and minimum values are simulation dependent and are given below. The proposal distribution from which a new sample  $Y$  is drawn is a normal distribution, that is  $Y \sim N(C_t, \sigma_p)$ , where  $N$  is a normal distribution of mean  $C_t$  and width  $\sigma_p$ .

Table I. Comparison of the performance of the IDL routine CURVEFIT against the simulated annealing routine. The true parameter values are listed in Section 3.1. Both SA and MCMC are limited to considering the range of values  $0.1 \leq A_1 \leq 2000$ ,  $-5.0 \leq c_1 \leq 5.0$ ,  $0.1 \leq \sigma_1 \leq 4.0$  and  $0.1 \leq \mu \leq 2000$ . Variables are quoted as  $\text{mean}[\sigma, \sigma^2(v)/\Upsilon(v)]$  where  $\sigma$  is the standard deviation of the distribution of results found.

variable	CURVEFIT	SA	MCMC
$A_1$	1434[25.4,1.50]	1434[26.6,1.51]	1436[30.2,1.43]
$c_1$	0.00[0.02,1.16]	0.00[0.01,1.17]	0.00[0.02,1.32]
$\sigma_1$	1.00[0.02,1.50]	1.00[0.02,1.52]	1.00[0.02,1.41]
$\mu_1$	114.0[4.04,1.19]	114.0[4.14,1.25]	114.0[4.3,1.28]

### 3. Results

#### 3.1. SINGLE EMISSION LINE

The first spectrum is based on the Fe XVI 360.76Å emission observed by SOHO-CDS in the execution of the SYNOP/v1 study. In this study, a window of emission from 360.123Å–361.386Å is observed using 18 pixels. The simulated line is centred in a window of 18 pixels, which has a constant background emission ( $\mu = 114$  photons/px), and  $A_1 = 1434$  photons/px,  $c_1 = 0.0$ ,  $\sigma_1 = 1.0$ . Simulated data  $D_i$  at pixel  $x_i$  is drawn from a Poisson distribution with mean  $F_i$ . Results from SA and MCMC are compared to those from the IDL program CURVEFIT; all three algorithms are seeded with the same initial estimate;  $\mu = \min(D)$ ,  $A = \max(D) - \min(D)$ ,  $c = \sum_{i=1}^{N_d} x(D_i - \mu) / \sum_{i=1}^{N_d} (D_i - \mu)$ ,  $\sigma^2 = \sum_{i=1}^{N_d} (D_i - \mu)(x - c)^2 / \sum_{i=1}^{N_d} (D_i - \mu)$ . Note that seeding with special values is not required for simulated annealing, as the routine moves randomly through the search space at large values of annealing temperature  $T$ , essentially rendering all starting points equivalent. The estimated values are however, crucial to the effective operation of the deterministic CURVEFIT routine that is used for comparison. Variable values in SA and MCMC as well as CURVEFIT are unconstrained.

Consider first the results of the analysis of a single line using SA compared to the IDL routine CURVEFIT. Table I shows that the answers found by SA are almost identical to those of CURVEFIT (the SA schedule is run with  $\delta = 0.91$  for 150 iterations and 100 trial configurations tried at each timestep; the weights  $W_i = 1, \forall i$ ).

Figure 3 shows performance diagnostics of the simulated annealing algorithm for the simulation. For the simulation (Figure 3(a)), the acceptance ratio begins high, indicating that the routine is performing like a random search through the entire search space. As the temper-



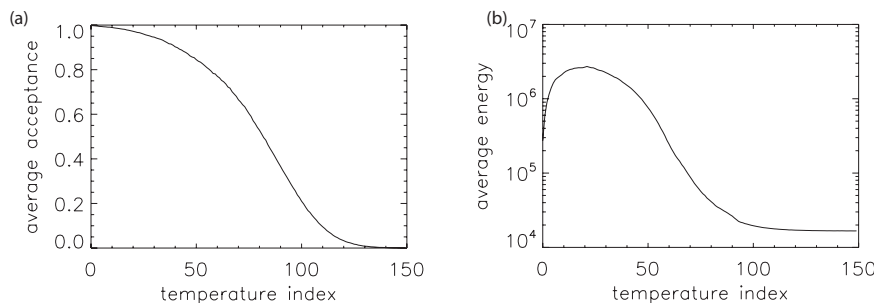


Figure 3. Average acceptance ratio (a) and average energy ( $E$ ) (b) over 1000 fits of the single Gaussian simulated data. The acceptance ratio is defined as the ratio of the accepted number of moves to a higher energy divided by the attempted number of moves to a higher energy.

ature cools, the routine moves more towards a more localized search, exploring only smaller increases in energy. Finally, the routine becomes a downhill search with effectively zero probability of an uphill move being acceptable. Figure 3(b) shows the energy behaviour. Note that the energy increases initially, indicating that the initial guess is not the worst possible one, and that the routine is acting like a random search, since it almost immediately jumps to a location in the search space which has a much higher energy. Despite this, SA still finds values equivalent to those found by CURVEFIT. The energy measure is reduced by a factor of at least ten from its maximum value (and less than the initial value), indicating that minimization of the energy function has occurred. MCMC has a comparable fitting performance to SA (the final 1000 draws of a total of 10 000 draws are used to calculate the values in Table I). Both MCMC and SA are performing at least as well as CURVEFIT in this simple situation. Note that MCMC errors correctly take into account the Poisson nature of the source data, whereas SA and CURVEFIT do not.

Both SA and MCMC are considerably slower than CURVEFIT, although in general the time penalty will vary depending on the problem being solved and the configuration of the algorithm (for example, number of cooling steps in SA or number of draws from the posterior distribution in MCMC). It is important to note, however, that the increased amount of time these algorithms take over CURVEFIT do bring two important features to line-fitting. Firstly, SA does away with the need to seed the fitting algorithm with a special initial value, eliminating user bias. Secondly, the results of the MCMC simulation properly take into account the full state of knowledge of the problem, including likely parameter values and the noise properties of the observation. This is especially important in low signal-to-noise ratio observations,

where the common assumption of Gaussian-distributed noise may not be applicable and other noise distributions, such as Poisson, have to be used. Example analyses are given below in Section 5 demonstrating the difference that inclusion of the proper noise distribution makes to the behavior of the fit in low-signal to noise regimes. In addition, the noise distribution in MCMC is not necessarily limited to Poisson or Gaussian distributions, and so MCMC can be used to fit lines where the noise is arbitrarily parameterized.

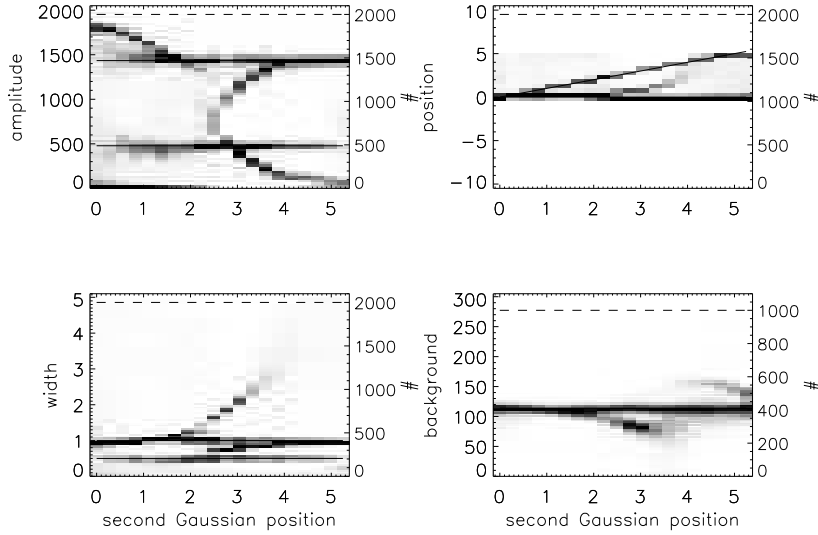
Finally, the summary statistics, such as error bars on the average values to the line parameters, properly reflect the complete knowledge of the model and the data, rather than the much more general, and less specific, knowledge implicit in algorithms such as CURVEFIT. For example, if the user has very good knowledge of the line width, this can be incorporated with an appropriate prior, and the resulting final answer for the line width will be more accurate, and more precisely determined. Further, all other parameters in the line fit will also be affected by this information, leading to a more correct determination of the line profile given the very good knowledge of the line width.

### 3.2. TWO EMISSION LINES

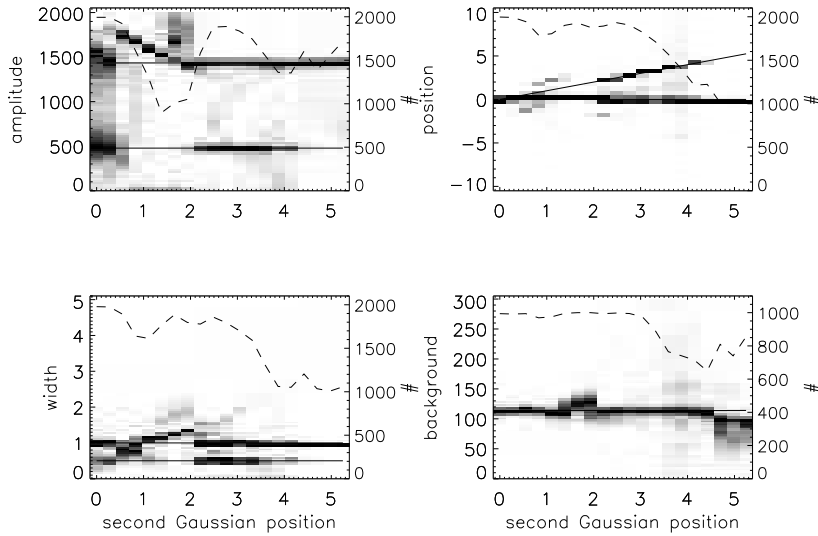
The second spectrum is based on the first with the inclusion of a second, smaller line, contaminating the larger line, which we will assume is the one we are interested in measuring. This spectrum can be used to judge the efficacy of these routines in correctly separating the two lines. The second line has amplitude  $A_2 = A_1/3$ , width  $\sigma_2 = \sigma_1/2$  and moves across the spectrum with a center  $c_2$ ,  $0 \leq c_2 \leq 5$ . Both SA and MCMC limit variable values to  $0.1 \leq A_1, A_2 \leq 2000$ ,  $-5.0 \leq c_1, c_2 \leq 5.0$ ,  $0.01 \leq \sigma_1, \sigma_2 \leq 5.0$  and  $0.1 \leq \mu \leq 2000$ .

Figures 4, 5, and 6 show that the inclusion of a second contaminating line is a serious problem for CURVEFIT, SA, and MCMC. All routines find that the true amplitude value is particularly difficult to find when  $0 \leq c_2 \leq 2$ . For Figure 5, variable values outside the y-axis range in the above plots are considered not acceptable and are therefore discarded.

All routines exhibit broadly similar behaviors for different positions  $c_2$ , but with some significant variation. CURVEFIT often does not find values within the acceptable range at all, indicating a significant breakdown of fitting in a large number of cases. The routine becomes trapped in regions of the search space distant from the location of the true minimum and cannot move from these local minima given the deterministic nature of the routine. Another significant problem, particularly for SA and CURVEFIT can be seen in the frequent occurrence of fits inside the permitted range, but well away from the true values. For  $2 \leq c_2 \leq 4$ , SA



*Figure 4.* Distributions of values found by Simulated Annealing for multiple locations of the smaller emission lines. Darker areas indicate a higher frequency of occurrence. Horizontal solid lines denote the true values of the emission spectrum parameters. The dashed line (with reference to the right hand axis values) indicates the number of valid values at every position of the second emission line. The annealing schedule was  $\delta = 0.91$  for 120 temperature reductions and 100 trials at each temperature.



*Figure 5.* Same as Figure 4, except using CURVEFIT to find the emission spectrum parameters. The dashed line (with reference to the right hand axis values) indicates the number of valid values at every position of the second emission line.

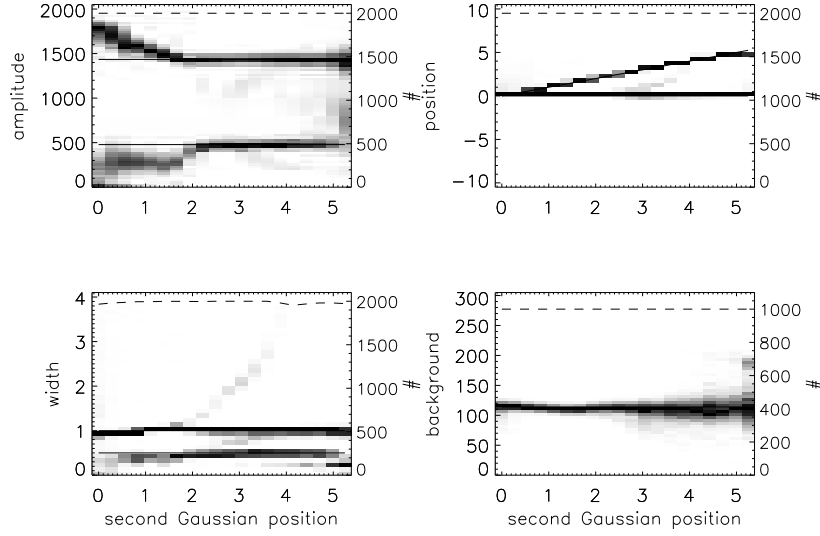


Figure 6. Distributions of values found by MCMC for multiple locations of the smaller emission lines. Darker areas indicate a higher frequency of occurrence. Horizontal solid lines denote the true values of the emission spectrum parameters. The dashed line (with reference to the right hand axis values) indicates the number of valid values at every position of the second emission line.

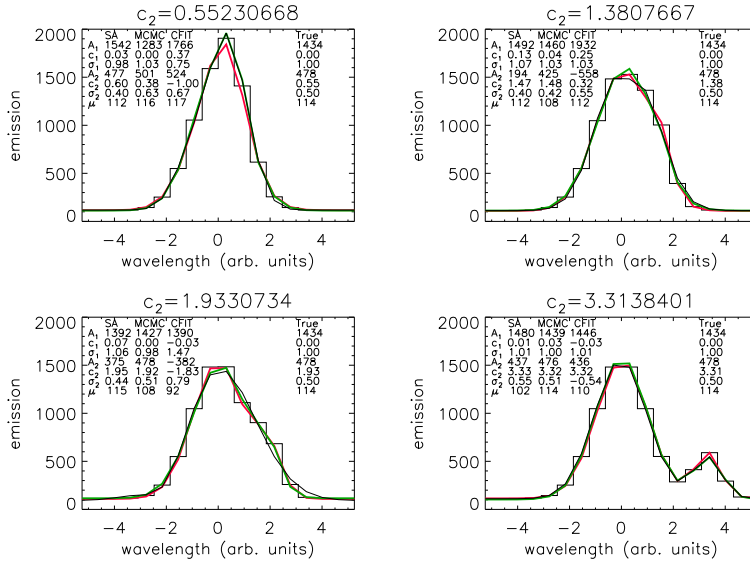


Figure 7. Example fits for each fitting algorithm for different positions ( $c_2$ ) of the smaller contaminating line. The true data is given as a histogram. CURVEFIT results are plotted in black, SA, green, and MCMC, red. Although the fits appear reasonable, it is clear from Figures 4, 5, and 6 that the parameter values may not be close to the true values.

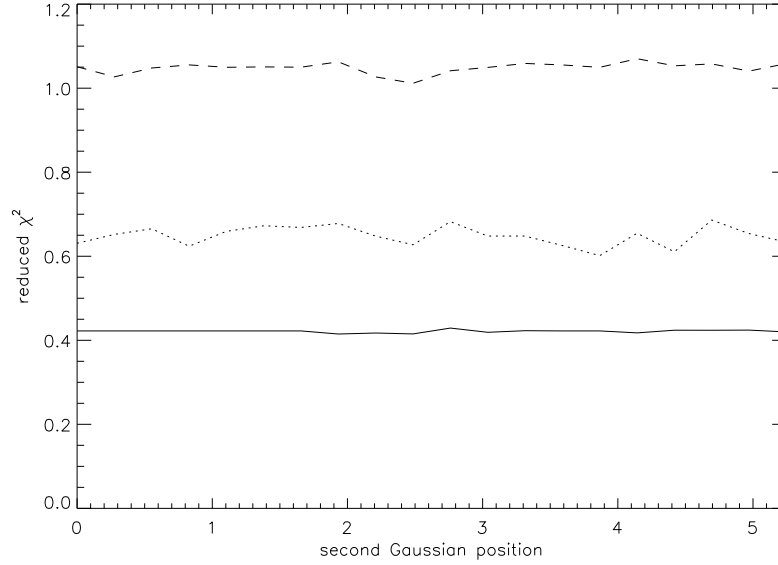


Figure 8. Average reduced  $\chi^2$ -squared measure for CURVEFIT (dotted), SA (dashed) and MCMC (solid).

and CURVEFIT fit the overall shape by minimizing  $E$  but do poorly at deriving the proper variable values, even though they find local minima – the information (contained in the true variable values) is smudged out by the noise. MCMC does considerably better in this range – it is not minimizing  $E$ , but rather finding those variable values that are most likely to give rise to the observed data given the spectral model. Hence the difference between SA and CURVEFIT and MCMC is largely due to the different approach to finding variable values.

The reduced  $\chi^2$  (Figure 8) values show that the routines are finding good fits to the data on average, in the sense that the sum of the squares of the differences between the data and model fit, weighted by the data values, are close to or below one. Again, MCMC out performs the other two methods by this measure also. Figure 9 shows that the average values found for the main Gaussian are closer to the true values over more of the range of second Gaussian positions.

#### 4. Application to Mg X / O IV spectrum in SOHO/CDS

The Mg X (625Å) line as observed in CDS can be contaminated by an overlapping O IV triplet (624.62Å, 625.13Å, 625.85Å), as well as background emission. However, as a consequence of their formation mechanism, the lines in the triplet are not independent and appear at

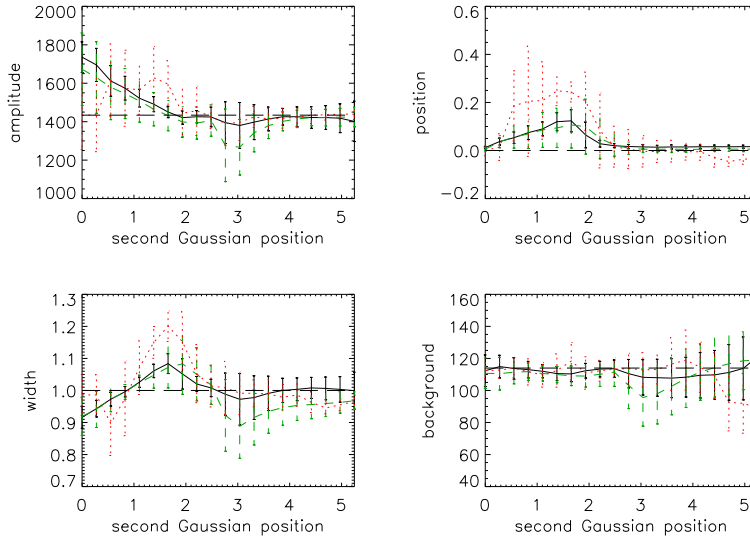


Figure 9. Comparison of the main Gaussian parameter values and background as found by SA (green, dashed line) MCMC (black, solid line) and CURVEFIT (red, dotted line), as a function of the second Gaussian position, with an error bar of plus/minus one standard deviation. True values are indicated by the dashed lines.

fixed offsets relative to each other, and have fixed relative amplitudes. Hence the entire spectrum can be represented by three variables for the Mg X line, a constant background value, and three variables representing any one of the O IV triplet. O IV 625.85Å is chosen as it has the largest amplitude and is most distant from the Mg X line; this makes this line the easiest to consider fitting out of the triplet. The model is

$$F_i = \mu + A_1 G(c_1, \sigma_1) + A_2 G(c_2, \sigma_2) + 0.670 A_2 G(c_2 - 0.723, \sigma_2) + 0.336 A_2 G(c_2 - 1.231, \sigma_2) \quad (6)$$

where  $G(c, \sigma) = \exp[0.5((x_i - c)/\sigma)^2]$ . A hybrid algorithm is used to determine values of the seven variables required. The hybrid algorithm is seeded with a fixed estimate which is passed to a SA algorithm (as was pointed out above, the initial seed matters little in this SA implementation). The best answer found after SA then seeds a MCMC algorithm. The post burn-in sample distribution is then used to calculate mean values and other parameters summarizing the distribution. A typical set of post burn-in distributions are shown in Figure 10; a set of summary measures of these distributions are shown in Table 4.

The advantage of this hybrid algorithm is that it permits a thorough and unbiased examination of the search space, and is guaranteed to find local minima without any user intervention. The errors associated with each value also correctly take into account the Poisson nature of the

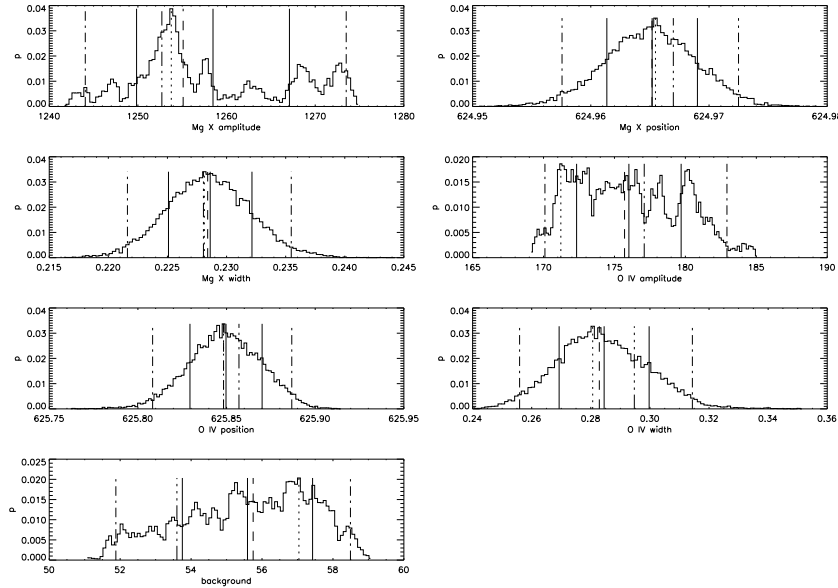


Figure 10. Distributions of values found after burn-in. Indicated on each distribution are mean values (thick solid) plus/minus the standard deviation (thin solid), mode (dotted), median (dashed), the 95% credible interval (dot-dashed lines), and values found interactively using MPCURVEFIT (dot-dot-dot-dashed).

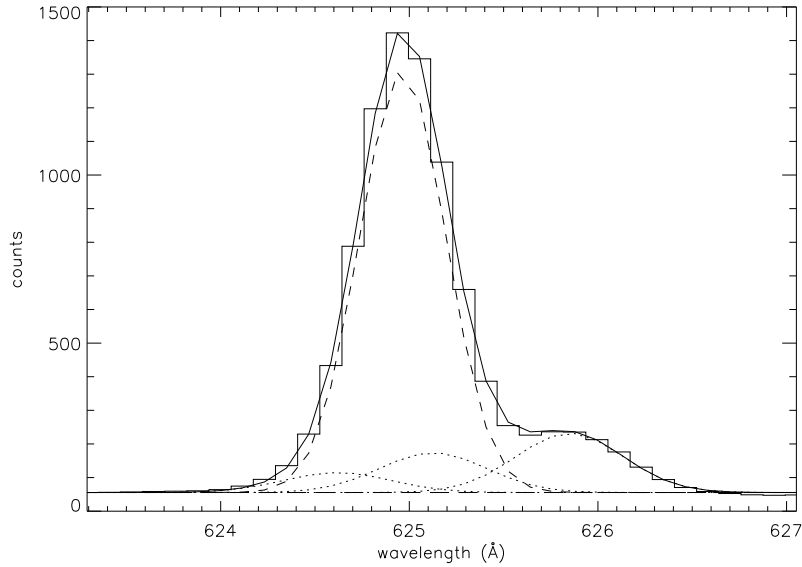


Figure 11. Original CDS data (solid line, histogram) overplotted with average fit (solid line, taken from mean values in Table 4). Also plotted is the Mg X line (short dashed line), the O IV triplet (dotted lines) and the background (long dashed line). The reduced  $\chi^2$  value of the MCMC fit is approximately 0.317; using the variable values found through MPCURVEFIT is approximately 0.381.

Table II. Results of fitting Equation (6) to data by the hybrid algorithm described in Section 4. Mean, standard deviation (s.d.), mode, median and credible interval values are calculated on the post burn-in distributions shown in Figure 10. The credible interval is measured by starting at the tails of the distribution and moving in until the percentage of values lying outside the limit is 2.5% .

variable	mean[s.d.]	mode	median	95% credible interval
Mg X $A_1$	1258.47[8.62]	1253.77	1255.10	1244.06, 1273.49
Mg X $c_1$	624.965[0.004]	624.965	624.965	624.957, 624.972
Mg X $\sigma_1$	0.229[0.004]	0.229	0.229	0.221, 0.235
O IV $A_2$	176.02[3.48]	171.22	170.090	170.10, 182.93
O IV $c_2$	625.85[0.023]	625.85	625.85	625.81, 625.89
O IV $\sigma_2$	0.284[0.015]	0.280	0.282	0.256, 0.314
$\mu$	55.59[1.83]	57.04	55.75	51.87, 58.49

noise. A CURVEFIT-based fit relies on multiple and inevitably biased initial choices, and possible user intervention should the routine crash. In addition, the errors are assumed to be normally distributed, which is a valid approximation only in the large-count limit. The hybrid routine inherently assumes Poisson statistics. Finally, it should be noted that the hybrid algorithm arrived at a good fit without any user intervention during the run, or any particular seed value to start the algorithm.

## 5. Application to Poor Signal-to-noise Ratio Data

For photon counting spectrometers, the assumption of Gaussian- distributed data is not valid when the number of counts is very low. Hence the assumptions implicit in least-squares algorithms such as CURVEFIT break down, leading to incorrect fits.

Table III compares the results of applying the hybrid algorithm to two simulated spectra with very low counts (with slightly varying prior knowledge) to results arising from applying CURVEFIT to the same data. Two simulated spectra are considered. The first has  $A_1 = 2$ ,  $c_1 = 0$ , and  $\sigma_1 = 1$ . The second spectrum has the same emission line, this time on a background of  $\mu_1 = 5$ . The signal to noise ratio of this second spectrum is less than one. Further, two different sets of prior information are used, labeled (a) and (b); variables are restricted to  $0 \leq A_1, \mu_1 \leq 10$  (Jeffrey's prior),  $-5 \leq c_1 \leq 5$  (uniform prior),  $0.1 \leq \sigma_1 \leq 4$  (Jeffrey's prior), as applicable, for prior set (a). Prior set (b) changes the range on the line width to  $0.8 \leq \sigma_1 \leq 1.2$ ; all other priors remain unchanged, and so changes in average parameter values



and errors are solely due to the effect of improving the information on the line width. The SA algorithm looks for minima with parameter values within the limits described above.

Simulated spectra are created by drawing from a joint Poisson distribution with mean equal to the true value of the spectrum. The hybrid algorithm is started with a uniform-random chosen initial value and seeds an SA algorithm. The best result found is then used to seed a MCMC algorithm and CURVEFIT. This process is repeated 1000 times, and the mean values and one standard deviation errors for each of the parameters are given in Table III.

As can be seen, the MCMC algorithm gives answers which are on average more accurate and more precise than either CURVEFIT or SA. This is a reflection of the fact that we have properly accounted for the noise in the system. The assumption of Gaussian-distributed errors leads to less accurate and less precise values in the case of a low signal.

Further, the use of better information (comparing results labeled *(a)* and results labeled *(b)*) in the form of different priors also improves the mean value and the error. This demonstrates that when fitting lines, the hybrid algorithm takes advantage of all the information available to the user in order to get the best possible estimates of the line parameter values and errors.

This strongly suggests that in low signal-to-noise ratio data, a full understanding of the noise properties of the data, and the inclusion of any other relevant information, is essential in understanding the emission spectrum and in obtaining the best possible estimates of parameter values and errors.

In higher signal-to-noise ratio spectra, Poisson-distributed noise can be approximated by Gaussian distributed noise with width equal to the square root of the number of counts observed, or, equivalently, by assuming a “Poisson weighting” in CURVEFIT. However, in such routines it is difficult to also include the effects of any extra knowledge, such as improved knowledge of the line width, or limits on amplitudes (see Figure 7).

## 6. Conclusions

Simulated annealing algorithms and Bayesian posterior distribution analysis have been successfully applied to the determination of spectral model parameter values<sup>1</sup> It is clear that disentangling two overlapping lines in the presence of noise is a difficult problem. The search space

---

<sup>1</sup> Software and examples are available for download URL <http://umbra.nascom.nasa.gov/people/ireland/linefit.html>.

Table III. Comparison of the performance of the IDL routine CURVEFIT against SA, and MCMC seeded with the best result from SA (the hybrid routine). Parameter values are quoted as (value) $\pm$ (standard deviation). The true parameter values are  $A_1 = 2$ ,  $c_1 = 0$ ,  $\sigma_1 = 1$ , and  $\mu_1 = 5$ . Simulation (1) refers to an emission line with no background; simulation (2) has a background. Labels (a) and (b) refer to different priors used. CURVEFIT parameters given in *italics* refer to average values and their concomitant one-standard-deviation errors when the answers returned by CURVEFIT lie within the range of values given by the priors. The number of parameter sets lying within these ranges are (1a) 977/1000, (1b) 273/1000 (2a) 842/1000 and (2b) 131/1000.

simulation	parameter	curvefit	SA	MCMC
(1a)	$A_1$	$3.63\pm 2.39$ , <i><math>3.44\pm 1.76</math></i>	$3.23\pm 1.62$	$2.22\pm 1.00$
	$c_1$	$-0.02\pm 0.52$ , <i><math>-0.01\pm 0.50</math></i>	$-0.02\pm 0.52$	$0.00\pm 0.39$
	$\sigma_1$	$0.66\pm 0.40$ , <i><math>0.67\pm 0.39</math></i>	$0.69\pm 0.40$	$1.09\pm 0.36$
(1b)	$A_1$	$3.60\pm 3.48$ , <i><math>2.25\pm 0.79</math></i>	$2.28\pm 0.87$	$2.02\pm 0.72$
	$c_1$	$0.01\pm 0.50$ , <i><math>0.02\pm 0.43</math></i>	$0.01\pm 0.49$	$0.00\pm 0.39$
	$\sigma_1$	$0.72\pm 0.41$ , <i><math>0.98\pm 0.11</math></i>	$0.91\pm 0.16$	$1.00\pm 0.06$
(2a)	$A_1$	$6.90\pm 4.99$ , <i><math>5.92\pm 2.75</math></i>	$5.47\pm 2.65$	$2.04\pm 1.45$
	$c_1$	$0.14\pm 4.76$ , <i><math>0.09\pm 1.87</math></i>	$0.10\pm 2.05$	$0.05\pm 1.07$
	$\sigma_1$	$1.80\pm 7.79$ , <i><math>0.77\pm 0.78</math></i>	$1.07\pm 1.19$	$1.27\pm 0.53$
	$\mu_1$	$3.56\pm 4.60$ , <i><math>4.68\pm 0.84</math></i>	$4.50\pm 1.12$	$4.95\pm 0.80$
(2b)	$A_1$	$9.43\pm 98.3$ , <i><math>3.51\pm 1.43</math></i>	$3.34\pm 1.48$	$1.48\pm 1.13$
	$c_1$	$0.03\pm 2.68$ , <i><math>0.10\pm 1.34</math></i>	$0.09\pm 2.08$	$0.02\pm 1.05$
	$\sigma_1$	$0.90\pm 1.73$ , <i><math>0.99\pm 0.11</math></i>	$0.92\pm 0.16$	$0.99\pm 0.04$
	$\mu_1$	$4.42\pm 2.71$ , <i><math>4.74\pm 0.59</math></i>	$4.77\pm 0.59$	$5.14\pm 0.58$

becomes filled with minima that are often distant from the location of the true values. Different search strategies locate different minima and so map out the search space.

One advantage these algorithms have over CURVEFIT and other deterministic routines is their stability when faced with noisy data – no matrices are inverted and so problems associated with ill-conditioned or singular matrices do not occur. Hence, these algorithms are guaranteed to find minima in all cases. Also, these algorithms can be easily forced to consider only reasonable areas in the search space.

MCMC presents the distinct advantage of both encoding the Poisson nature of the counting statistics (or any other kind of statistics, if the count distribution statistics are known) and allowing the generation of uncertainty measures (for example, standard deviations) from the post burn-in distribution of samples. This is a very powerful advantage, as it permits the analysis of very low emission count data (on either high or low backgrounds), taking complete account of the count statistics and the model, and leading to better parameter estimates than traditional

fitting routines. Summaries of uncertainty (such as standard deviations and 95% credible intervals) in the average values also properly incorporate the count statistics, and all available knowledge of the line.

It is envisaged that these techniques and related search algorithms (for example, genetic algorithms, McIntosh *et. al.* 1998) could be used to provide an initial input for more traditional routines, on both the number of emission features present and their characteristics. In particular, such analysis methods will be useful in the analysis of noisy, low-signal-to-noise line emission data from instruments on board the SOHO spacecraft, such as CDS and SUMER, and from EIS onboard Hinode.

### Acknowledgements

The author would like to thank Peter Young (RAL) for supplying the Mg X CDS data, and to C. Alex Young and Danielle Bewsher for their helpful suggestions. This work was supported by a NASA Science Research and Technology contract, NNH04CC31C.

### References

- Aarts, E. and Korst, J., 1990, Simulated Annealing and Boltzmann Machines, John Wiley and Sons, Chichester.
- Beveridge, C., 2006, Sol. Phys., *236*, 41.
- Brosius, J.W., Davila, J.M., Thomas, R.J., Monsignori-Fossi, B.C. 1996, Astrophys. Suppl. Ser., *106*, 143 (B96).
- Brynildsen, N., 1994, Profile fitting to CDS/SUMER data, CDS Software Note v1.0, No.21, Institute of Theoretical Astrophysics, University of Oslo, P. O. Box 1029 Blindern, N-0315 Oslo, Norway.
- Corana, A., Marchesi, M., Martini, C., Ridella, S., 1987, AMS. Trans. Math. Soft., *13*(3), 262.
- Davis, L. (ed.), 1987, Genetic algorithms and simulated annealing (Research Notes in Artificial Intelligence), Pitman/Morgan Kaufmann Publishers Inc., London.
- Gregory, P. C., 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support, Cambridge University Press, Cambridge.
- Harrison, R.A. Sawyer, E.C., Carter, M.K., Cruise, A.M., Cutler, R.M., Fludra, A., Hayes, R.W., Kent, B.J., Lang, J., Parker, D.J., Payne, J., Pike, C.D., Peskett, S.C., Richards, A.G., Culhane, J.L., Norman, K., Breeveld, A.A., Breeveld, E.R., Al Janabi, K.F., McCalden, A.J., Parkinson, J.H., Self, D.G., Thomas, P.D., Poland, A.I., Thomas, R.J., Thompson, W.T., Kjeldseth-Moe, O., Brekke, P., Karud, J., Maltby, P., Aschenbach, B., Bräuninger, H., Kühne, M., Hollandt, J., Siegmund, O.H.W., Huber, M.C.E., Gabriel, A.H., Mason, H.E., Bromage, B.J.I., 1995, Sol. Phys., *163*, 233.
- Ireland, J., 2005, Astrophys. J., *620*, 1139.

- Jeffrey, W., Rosner, R., 1986, *Astrophys. J.*, *310*, 473, 1986.
- Kashyap, V., Drake, J.,J., 1998, *Astrophys. J.*, *503*, 450.
- Kirkpatrick, S., Gelatt, C.D., Jr., Vecchi, M.P., 1983, *Science*, *220*(4598), 671.
- van Laarhoven, P. J. M., and Aarts, E., 1987, *Simulated Annealing: Theory and Applications*, R. Deidel Publishing Company, Kluwer Academic Publishers, Norwell.
- McIntosh, S W., Diver, D.A., Judge, P.G., Charbonneau, P., Ireland, J., Brown, J.C., 1998, *Astron. Astrophys. Supp. Ser.*, *132*, 145.
- Metropolis, N., Rosenbluth, A W., Rosenbluth, M N., Teller, A.H., Teller, E., 1953, *J. Chem. Phys.*, *21*(6), 1087.
- van Dyk, D.A., Connors, A., Kashyap, V.L., Siemiginowska, A., 2001, *Astrophys. J.*, *548*, 224.
- de Vicente, J., Lanchares, J., Hermida, R., *Phys. Lett. A.*, *317*, 415, 2003.
- Siarry, P., Berthiau, G., Durbin, F., Haussy, J., 1997, *ACM Trans. Math. Software*, *23*(2), 209.